Systems Biology and Computation

Felice Lightstone (LLNL) and Ian Korf (UC Davis)

LLNL-PRES-472611
LLNL-PRES-471711
LLNL-PRES- 472311
This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under contract DE-AC52-07NA27344. Lawrence Livermore National Security, LLC

LLNL-PRES-509691

What is the goal of Computational and Systems Biology?

The goal of computational and systems biology is to apply large-scale computational methods to the study of living systems at all scales, i.e. molecular, cellular, or organism.

The development and application of data-analytical and theoretical methods, mathematical modeling and physics-based computational simulation techniques is used to enhance the investigation and better understand the structure, function and dynamics of complex biological systems .

What is all the fuss about sequencing?

- Illumina HiSeq produces ~30 Gbp for ~$3,000 in 1 lane (there are 8 lanes)
- 10x coverage of the human genome in 1 lane!
- Bar-coding allows one to mix several experiments in a single sequencing reaction
- Sequencing costs will continue to drop
- $1000 genome (or less)

Sequencing and Cancer

- Every person is unique
- Every cancer is unique
- The best treatment depends on the specific person and the specific cancer
- In the future, we can sequence the genome and transcriptome of the patient and their cancer
  - SNPs (single nucleotide polymorphisms)
  - CNVs (copy number variants)
  - structural variants

Nothing in biology makes sense except in the light of evolution

- Nothing in personalized medicine makes sense except in the light of variation
- How are you going to use personal/cancer variation in your research and patient care?
- Does your knowledge/treatment take into account personal/cancer variation?
  - Why not?
  - What would it take to make this happen?

Genome analysis is not a solved problem

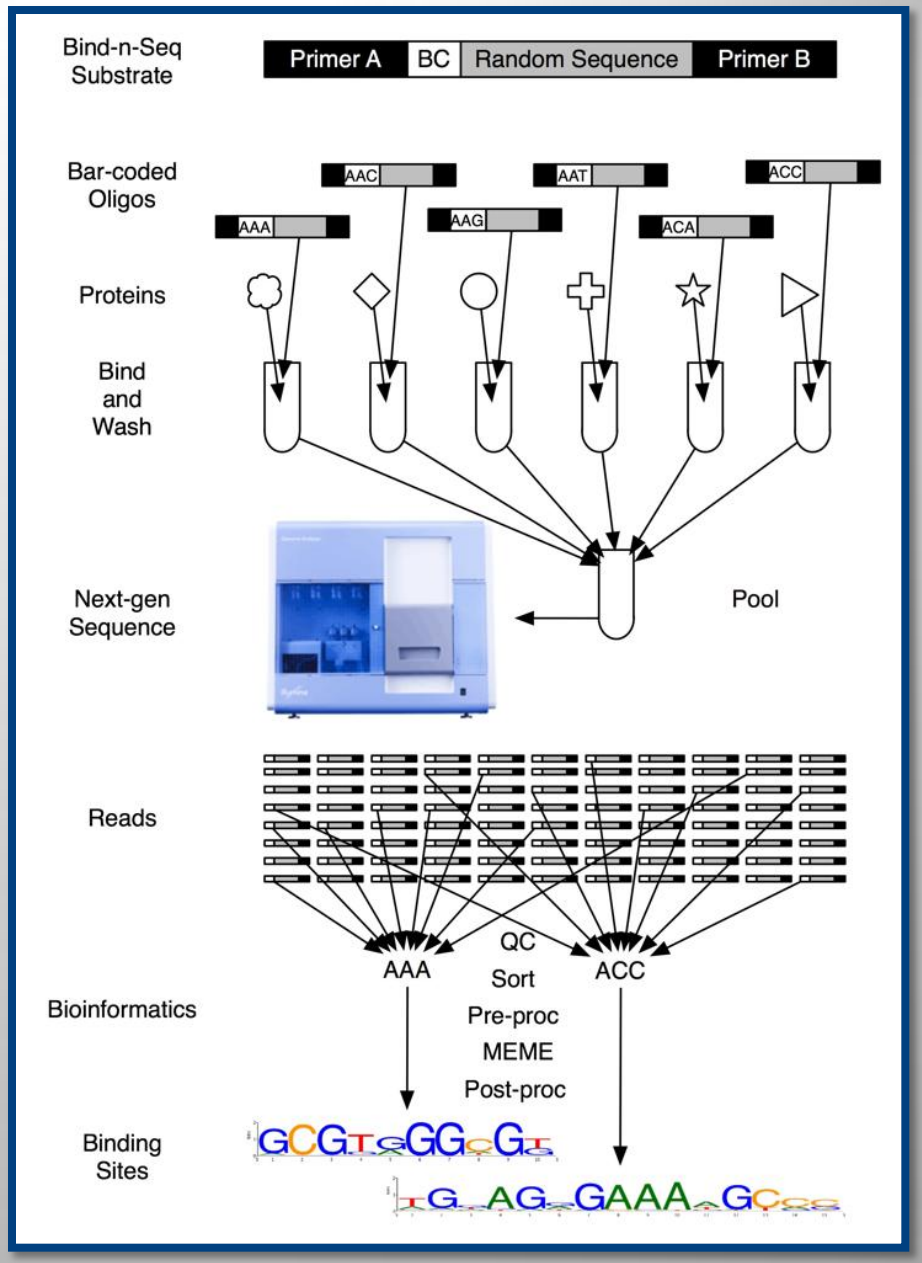
- Statistical issues
  - high dimensionality
  - not many samples in the early days
- Technical issues
  - Storage, retrieval
  - Privacy, security
- Computing issues
  - High-performance computing: GPU, FPGA, ASIC
  - Cloud computing

Sequencing can transform the way you work

- Sequencing has replaced microarrays (mostly)
  - More accurate, more data points, no hybridization
- Sequencing will replace qPCR
  - Why assay a dozen genes with qPCR when you can assay ALL of them with RNA-seq?
- New uses for sequencing technologies continue to be developed

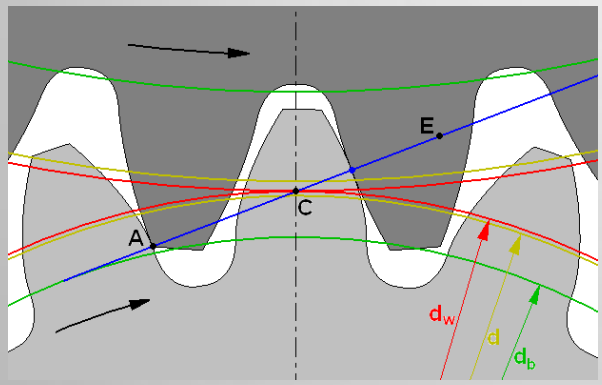
Example: Bind-n-Seq

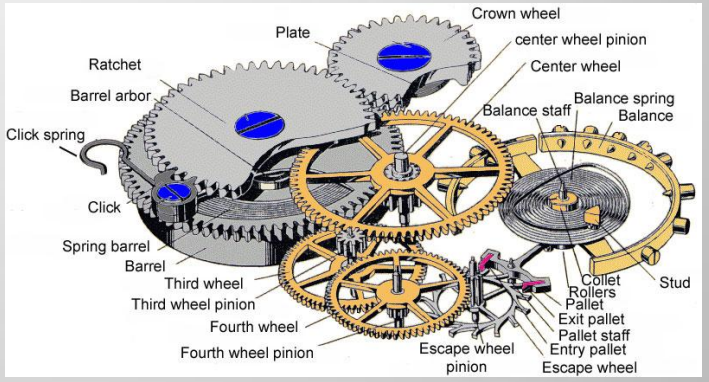
Biochemistry with a sequence read-out



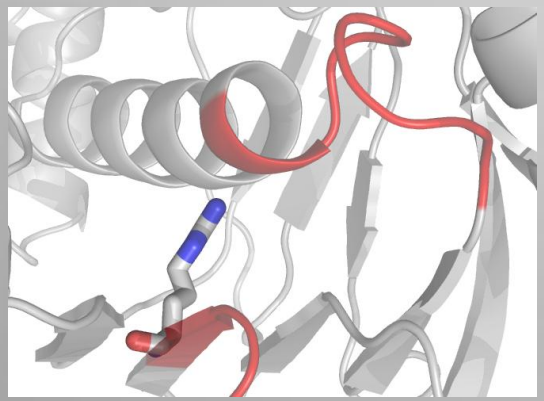
Systems Biology provides the whole-organism context

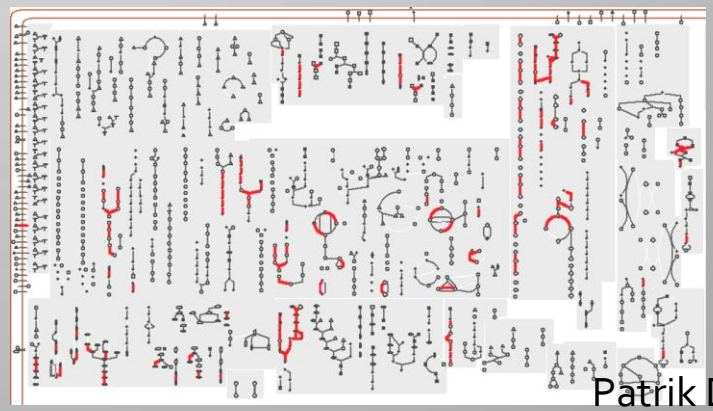
Proteins are the "gears" of a complex cellular machine. Systems Biology deals with how these gears interact, and what their role is in the functioning of the whole machinery.





<u>Multiscale</u> problem: from the molecular to the organismal level, genotype to phenotype



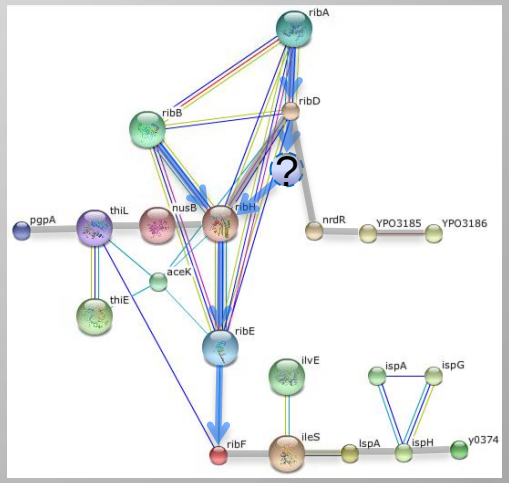


Patrik D'haeseleer

Systems Biology allows us to study protein function *within the context* of the overall cellular system

- Systems Biology **integrates all available sequence and structure based functional annotations**
- **Highlights gaps in our knowledge**, that more intensive structure-based methods can fill (e.g. investigate substrate specificity of key enzyme)
- **Identifies metabolically essential, and virulence related pathways as high-priority targets** for further investigation
- Protein association networks can be a **guide to identify missing or unexpected pathway members,** infer functional annotation, outline putative protein complexes, etc.
- Integrate from **molecular to phenotype** level

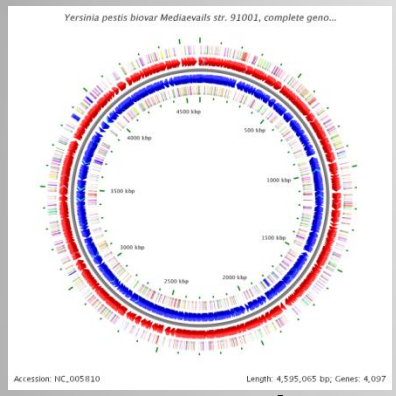
Complete metabolic pathway map, with essential pathways highlighted



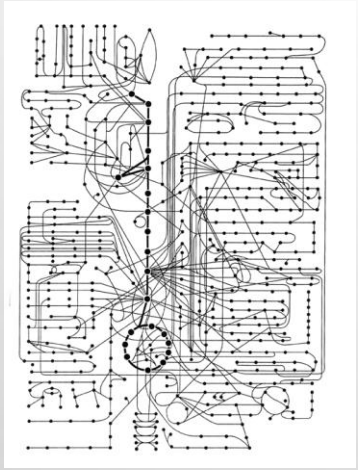
Network of metabolic pathway, operons, protein associations, and predicted protein interactions

Patrik D'haeseleer

Flux Balance Analysis is a systems biology method for assessing the metabolic capabilities of a genotype



Annotated genome



Network reconstruction

$$\frac{dM2}{dt} = 1 * V1 - 1 * V3 - 1 * V4 = 0$$

Stoichiometric matrix

Flux vector

R1 R2 ⋯ RN

M1 S11 S12

M2 S21 S22

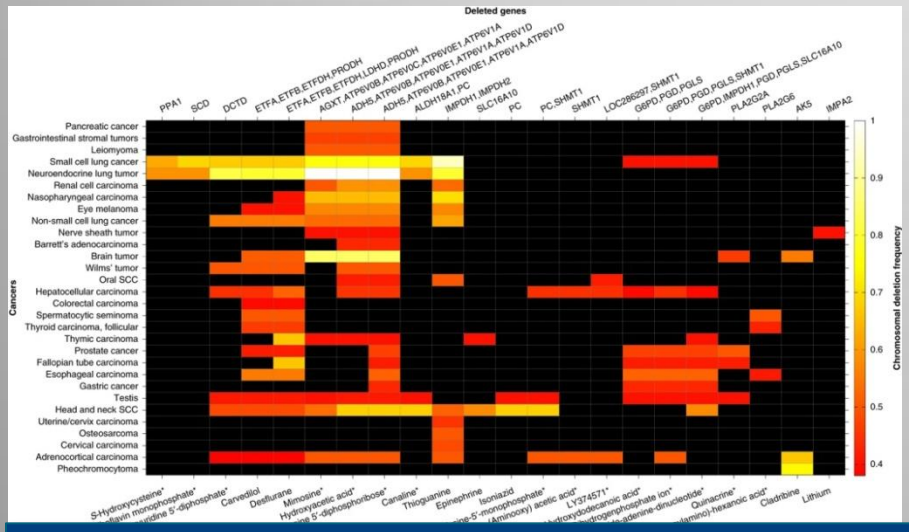
M5

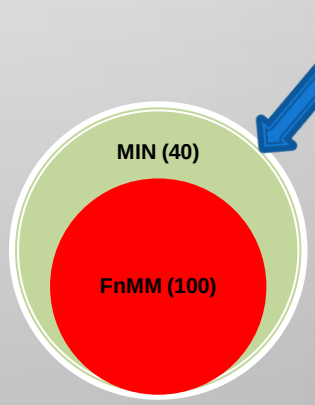
V1 V2 ⋮ = 0

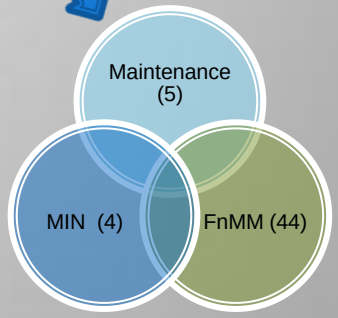
Constraints & Optimization (e.g. growth)

↓

Predicted metabolic phenotypes







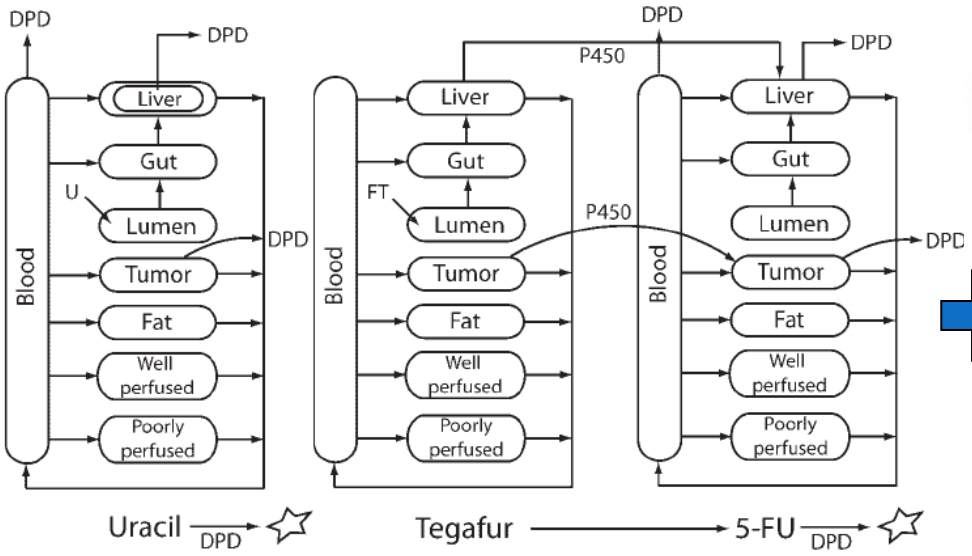
Critical single gene knockouts **Synthetic lethal mutations**

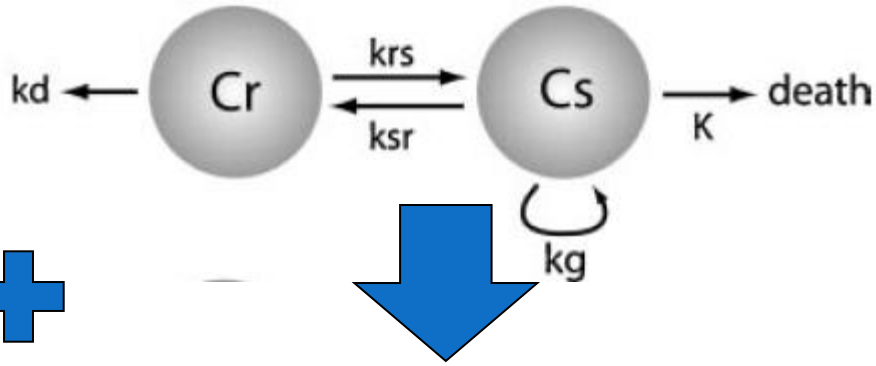
Candidate drug targets

Predicting selective drug targets in cancer through metabolic networks, Folger et al. 2011

Ali Navid

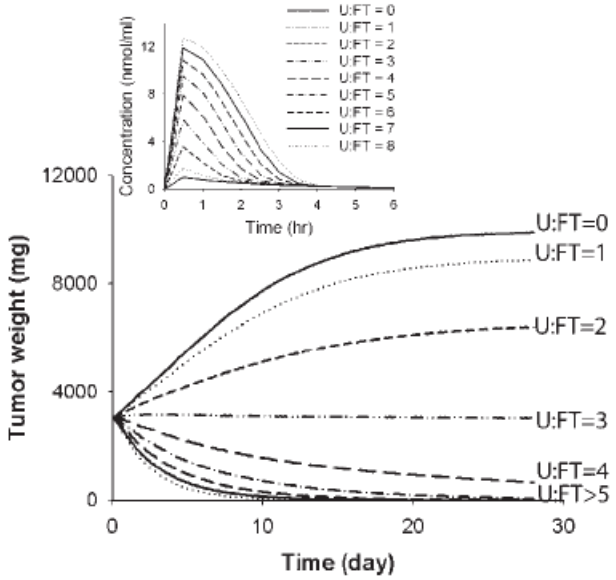
PK-PD modeling is a systems approach for predicting drug ADME and pharmacological outcome





Incorporate key input parameters from different sources to:

- **Estimate pharmacokinetic parameters**
- **Assess toxicological risks**
- **Design optimal therapeutic regimes**



Sung et al. (2009), J. Pharm. Sci., **98**, 1885

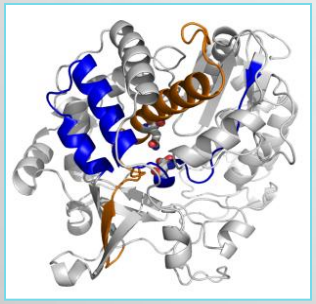
Ali Navid

Structure- and sequence-based tools are used to identify functionally critical residues and possible mutation points

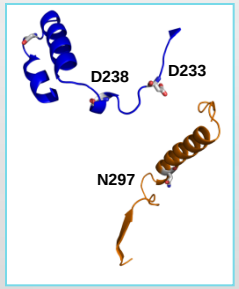
Protein sequence

```
>RNApolym NP_041277.1 [Poliovirus]
GEIQWMRPSKEVGYPIINAPSKTKLEPSAFHYVFE
GVKEPAVLTKNDPRLKTDFEEAIFSKYVGNKITEV
DEYMKEAVDHYAGQLMSLDINTEQMCLEDAMYGTD
GLEALDLSTSAGYPYVAMGKKKRDILNKQTRDTKE
MQKLLDTYGINLPLVTYVKDELRSKTKVEQGKSRL
IEASSLNDSVAMRMAFGNLYAAFHKNPGVITGSAV
GCDPDLFWSKIPVLMEEKLFAFDYTGYDASLSPAW
FEALKMVLEKIGFGDRVDYIDYLNHSHHLYKNKTY
CVKGGMPSGCSGTSIFNSMINNLIIRTLLLKTYKG
IDLDHLKMIAYGDDVIASYPHEVDASLLAQSGKDY
GLTMTPADKSATFETVTWENVTFLKRFFRADEKYP
FLIHPVMPMKEIHESIRWTKDPRNTQDHVRSLCLL
AWHNGEEEYNKFLAKIRSVPIGRALLLPEYSTLYR
RWLDSF
```

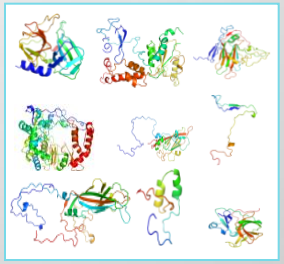
Structural model



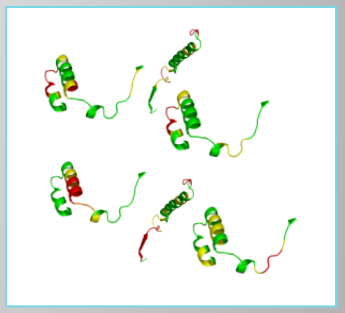
Structural fragments



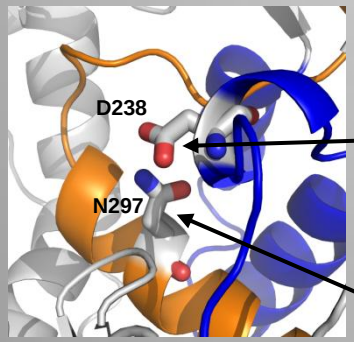
Structure database



Similar fragments detected



Example of identified functional



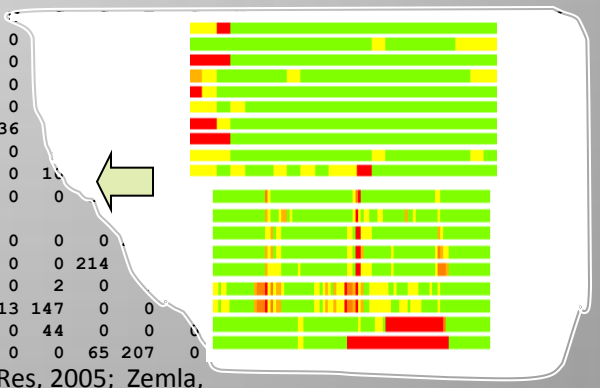
Sequence variability allowed in conserved regions

AA	Res	score	variety
D	233	100.0	D
Y	234	23.0	TYVL
T	235	9.6	RSTKAE
G	236	8.8	CDGANRKQX
Y	237	10.7	FHWYG
D	238	99.6	DE
A	239	11.5	STAG
S	240	11.7	TFSNQR
T	293	100.0	T
S	294	78.4	SDKYF
I	295	15.0	CLIAQMVTG
F	296	11.0	GMFIWAELVXN
N	297	81.3	NGH
S	298	23.8	TSH

Position specific matrices from structure conserved regions

AA	Res	A	V	L	I	P	M	F	
D	233	0	0	0	0	0	0	0	0
Y	234	0	48	1	0	0	0	0	0
T	235	5	0	0	0	0	0	0	0
G	236	16	0	0	0	0	0	0	0
Y	237	0	0	0	0	0	0	152	36
D	238	0	0	0	0	0	0	0	0
A	239	30	0	0	0	0	0	0	0
S	240	0	0	0	0	0	0	44	0
T	293	0	0	0	0	0	0	0	0
S	294	0	0	0	0	0	0	1	0
I	295	19	7	44	41	0	12	0	0
F	296	10	3	4	14	0	42	30	13
N	297	0	0	0	0	0	0	0	0
S	298	0	0	0	0	0	0	0	0

Structural alignments conserved regions colored in green

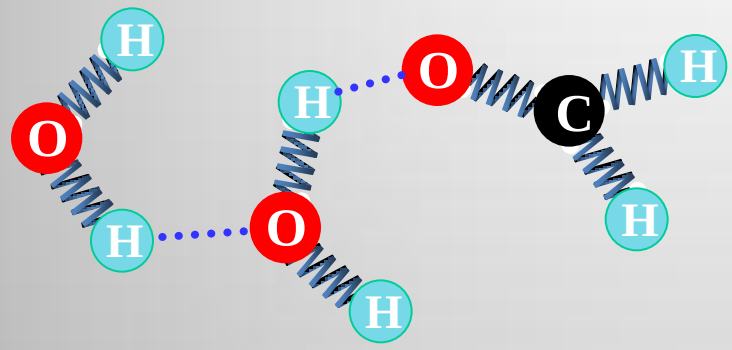


Zemla, Nucleic Acids Res, 2003; Zemla, Zhou, Slezak, Kuczmarski, Rama, Torres, Sawicka, Barsky, Nucleic Acids Res, 2005; Zemla, Geisbrecht, Smith, Lam, Kirkpatrick, Wagner, Slezak, Zhou, Nucleic Acids Res, 2007; Chakicherla, Zhou, Dang, Rodriguez, Hansen, Zemla, PLoS, 2009; Zemla, Lang, Kostova, Andino, Zhou, BMC Bioinformatics, 2011

Adam Zemla

Physics-based methods provide fundamental understanding of molecular systems

Empirically derived classical force field

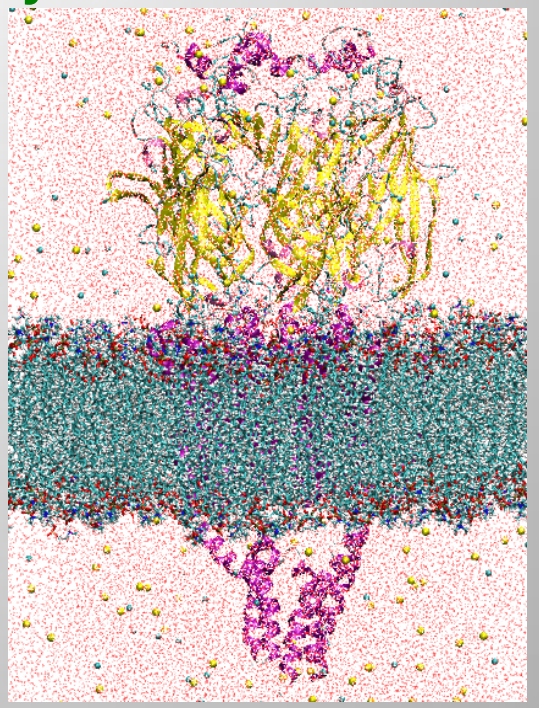


Use these forces to integrate the motion of each atom

$$a(t)=F(t)/m$$

$$r(t+\delta t)=2r(t)-r(t-\delta t)+a(t)\delta t^2$$

Routine Molecular Dynamics Simulations



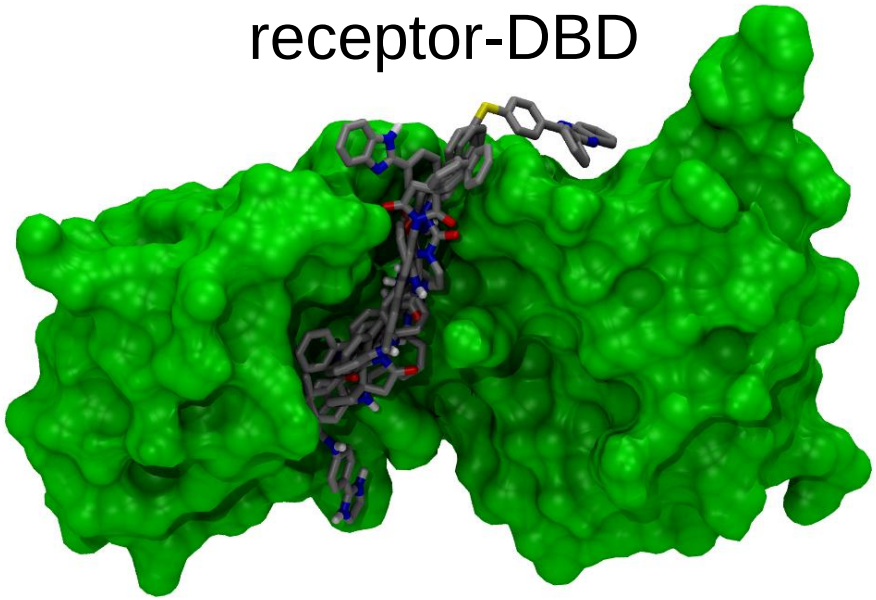
20 ns for ~250,000 atoms
72 hours on 512 Opteron processors

Questions this can address:

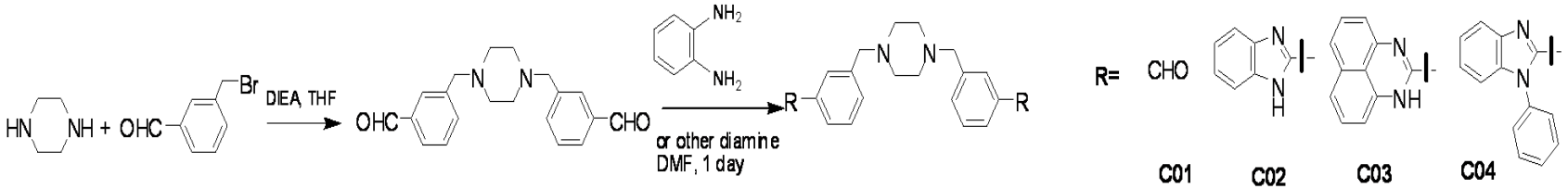
- What are the accessible conformations of this system?
- What contacts mediate a small molecule/macromolecular interaction?
- What are likely consequences of small molecule binding?

Molecular mechanics methods are used to design and develop small molecule inhibitors to therapeutic target

Seven candidate compounds docked to androgen receptor-DBD



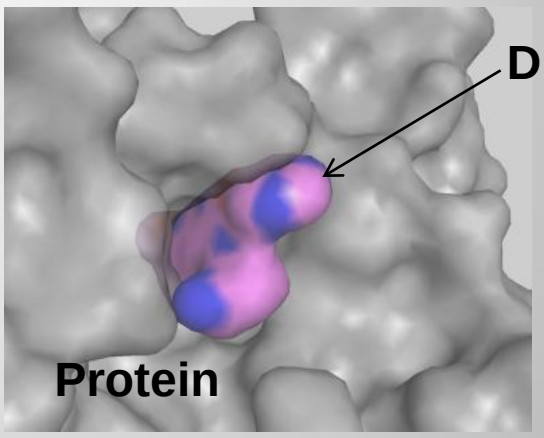
Compounds and analogs were synthesized



Dr. Ruiwu Li (UCDCC) and Brian Bennion (LLNL)

A variety of computational methods can help shorten the design and development time of novel antibiotics

- **High performance computing enabled designs of novel antibiotics against bio-defense pathogens**
  - New design is broad spectrum and predicted to target multi-drug resistant pathogens
- **LLNL-Trius team found and redesigned more efficacious drug candidate in 3 months (Stage 1 normally takes 2–5 years)**



Computational Screening:

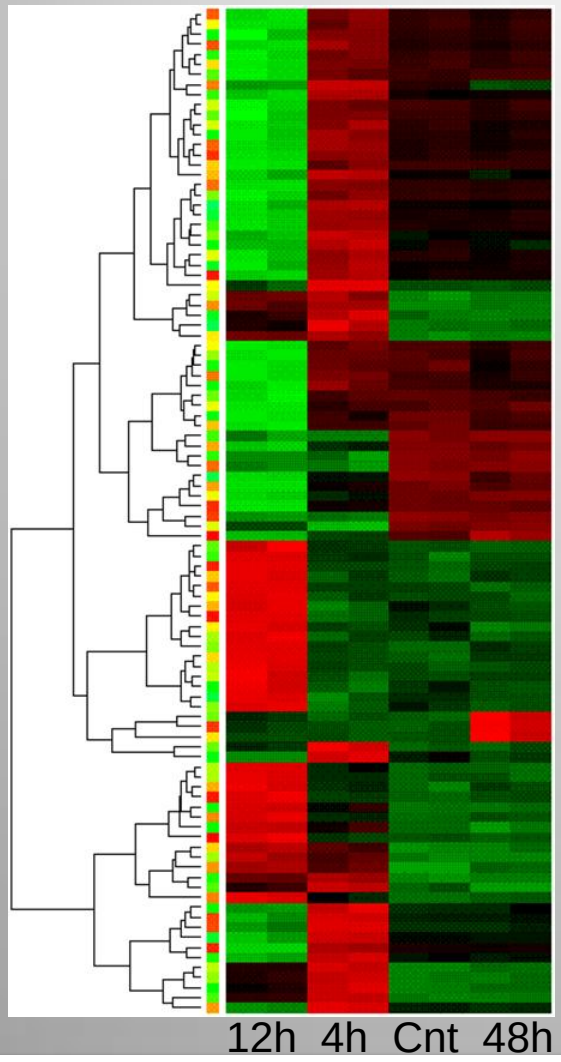
- Screened 8 million commercially available compounds
- Created and screened 10 x 60,000 compound virtual libraries
- Performed >5000 physics-based all- atom simulations

Integrated computational and experimental effort developed 8 compounds to move to preclinical trials

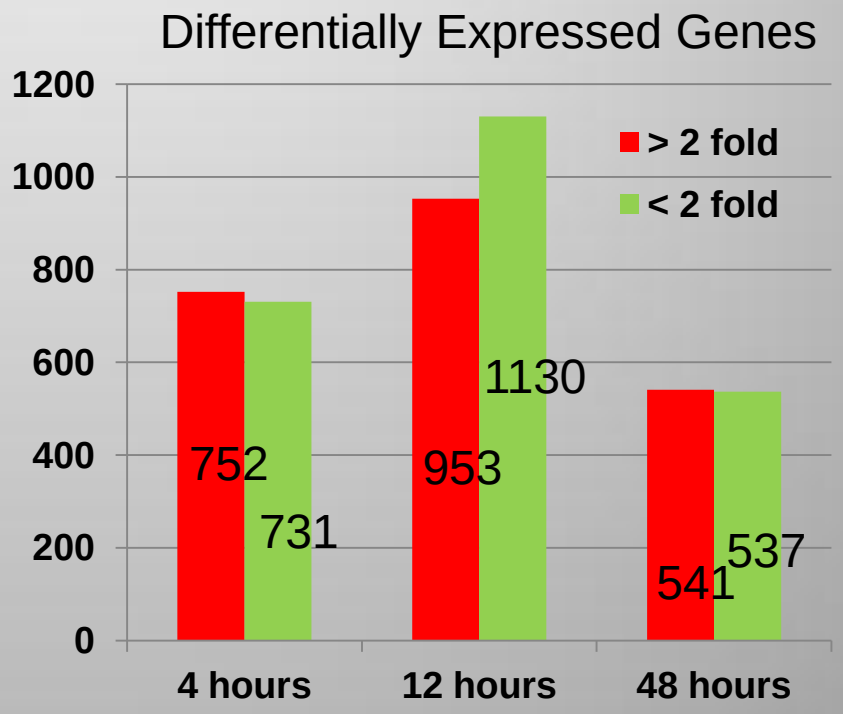
Felice Lightstone, Toan Nguyen, Sergio Wong, Ken Turteltaub, Mike Malfatti, Paul Jackson

The key to good models is good data

Osteolytic PC3 cells co-cultured with osteoblasts



More than 4.5K genes change in response to the Osteoblastic environment within 48 hr exposure



Gaby Loots

Contributors

- Patrik D'Haeseleer
- Ali Navid
- Adam Zemla
- Brian Bennion
- Ruiwu Li
- Gaby Loots